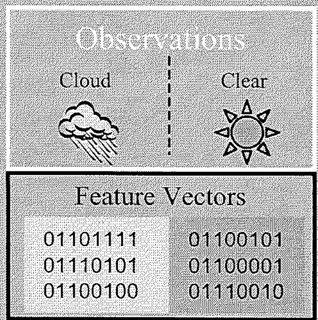
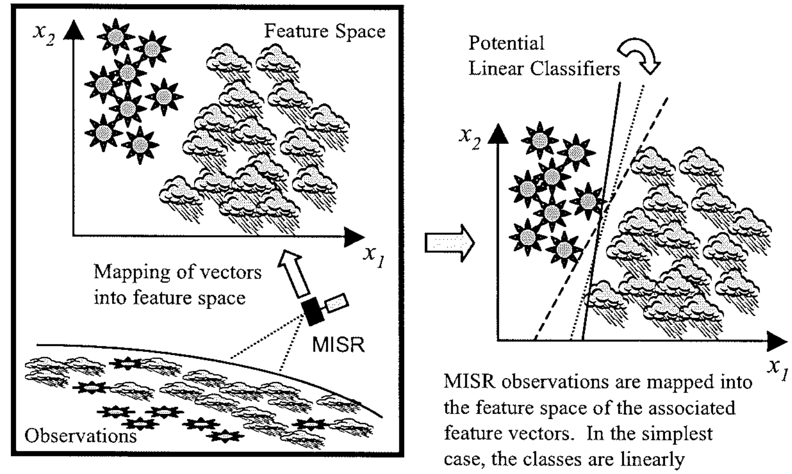
Support Vector Machines

Support Vector Machines (SVMs) are a type of supervised learning algorithm, other examples of which are Artificial Neural Networks (ANNs), Decision Trees, and Naïve Bayesian Classifiers. Supervised learning algorithms are used to classify objects into one of two or more categories, using training data consisting of objects labeled by a "supervisor" – typically a human "expert". SVMs are particularly adept at generalizing and they have been shown to achieve higher accuracy than other supervised learning algorithms on a benchmark character recognition test [DeCoste, D., and B. Schölkopf, Training invariant support vector machines, *Machine Learning, 46*, 161-190, 2002].



MISR observations can be represented mathematically as feature vectors containing the radiances of each pixel as measured by MISR's four color bands and nine cameras, along with contextual information.

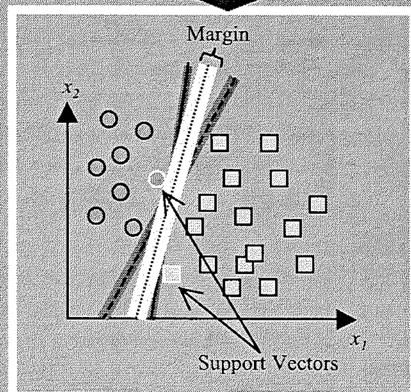
SVMs have been used to classify observations from the Multi-angle Imaging SpectroRadiometer (MISR) instrument. The objects the SVM attempts to classify are pixels from a MISR image, which may be classified as either "cloudy" or "clear". Instead of working with the observations themselves, the SVM utilizes the feature vectors obtained by combining a variety of MISR data. "Cloudy" and "clear" labels associated with these vectors are obtained by human inspection of the satellite scenes, or from existing MISR cloud classification algorithms.



MISR observations are mapped into the feature space of the associated feature vectors. In the simplest case, the classes are linearly separable, as shown on the right.

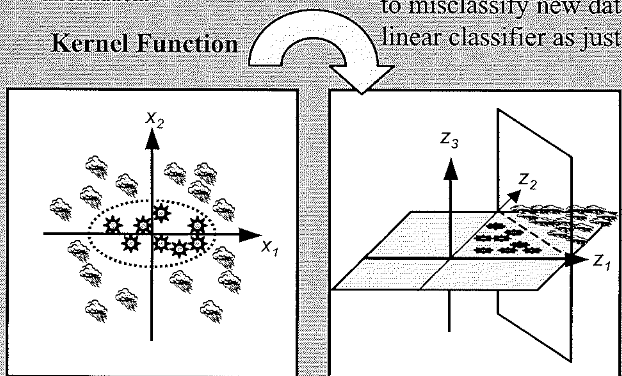
However, a number of potential linear classifiers exist and the problem is to choose the one which is most likely to be able to correctly classify additional data.

In the simplest case, a SVM finds the classifier that *best* separates the training vectors into two classes, choosing the one that maximizes the *margin*, the distance between the linear classifier and the nearest vector on either side. Intuitively, maximizing the margin means that the SVM is the most general classifier possible and the least likely to misclassify new data. It turns out it is computationally efficient to represent the best linear classifier as just the weighted sum of the vectors which lie on the margin. These vectors are known as *support vectors*.



Schematic showing the classifiers and the margins for the example above. The maximum margin is indicated by the yellow line, and the support vectors for this classifier are outlined in white.

Kernel Function



However, most classification problems aren't linearly separable, so SVMs map the feature vectors into a higher-dimensional space, as shown on the left. The optimal hyperplane that separates the example vectors in this higher-dimensional space corresponds to a nonlinear boundary in the original feature space, and through the use of a clever mathematical trick called a kernel function, this hyperplane can be found without ever computing the feature vectors in the higher-dimensional space.

MISR

CALIFORNIA INSTITUTE OF TECHNOLOGY 1891

JPL